UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/954,979	09/17/2001	Jon Rong-Wei Yi	01997-294001	7698

26161 7590 07/25/2005

FISH & RICHARDSON PC
P.O. BOX 1022
MINNEAPOLIS, MN 55440-1022

EXAMINER	
VO, HUYEN X	
ART UNIT	**PAPER NUMBER**
2655	

DATE MAILED: 07/25/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	09/954,979	YI ET AL.	
	Examiner	Art Unit	
	Huyen X. Vo	2655	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *10 February 2005*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-18* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-18* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *17 September 2001* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application (PTO-152)
6) ☐ Other: _____.

DETAILED ACTION

1. Applicant's request for reconsideration of the finality of the rejection of the last Office action is persuasive and, therefore, the finality of that action is withdrawn.

Claim Objections

2. Claims 12-13 are objected to because of the following informalities: there is a lack of antecedent basis. Claims 12-13 should not depend on claim 10, but rather depend on claim 11. Examiner treated claims 12-13 being dependent upon claim 11. Appropriate correction is required.

Claim Rejections - 35 USC § 102

3. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless – (b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

4. Claims 1-10 and 18 are rejected under 35 U.S.C. 102(b) as being anticipated by Hunt et al. (IEEE Publication).

5. Regarding claims 1 and 18, Hunt et al. disclose a method and a software stored on a computer-readable medium for selecting segments from a corpus of source utterances for synthesizing a target utterance, comprising: searching a graph in which each path through the graph identifies a sequence of segments of the corpus of source

utterances and a corresponding sequence of unit labels that characterizes a pronunciation of a concatenation of that sequence of segments, each path being associated with a numerical score that characterizes a quality of the sequence of segment (*sections 2.1-2.2 on pages 374-375*); wherein searching the graph includes matching a pronunciation of the target utterance to paths through the graph, and selecting segments for synthesizing the target utterance based on numerical scores of matching paths through the graph (*sections 2.1-2.2 on pages 374-375, Viterbi search algorithm propagates through the graph and picks the best paths*).

6. Regarding claims 2-3 and 5, Hunt et al. further disclose the method of claim 1 wherein selecting segments for synthesizing the target utterance includes identifying a path through the graph that matches the pronunciation of the target utterance and selecting the sequence of segments that is identified by the determined path (*sections 2.1-2.2 on pages 374-375, one best path is selected based on "concatenation cost"*), wherein determining the path includes determining a best scoring path through the graph (*sections 2.1-2.2 on pages 374-375, one best path is selected based on "concatenation cost"*), and concatenating the selected sequence of segments to form a waveform representation of the target utterance (*sections 2.1-2.2 on pages 374-375*).

7. Regarding claims 6-8, Hunt et al. further disclose the method of claim 1 wherein selecting the segments for synthesizing the target utterance includes determining a plurality of paths through the graph that each matches the representation of the

pronunciation of the target utterance (*sections 2.1-2.2 on pages 374-375*), wherein selecting the segments further includes forming a plurality of sequences of segments, each associated with a different one of the plurality of paths (*sections 2.1-2.2 on pages 374-375, inherent in Viterbi search algorithm*), and wherein selecting the segments further includes selecting one of the sequences of segments based on characteristics of those sequences of segments not determined by the corresponding sequences of unit labels associated with those sequences (*sections 2.1-2.2 on pages 374-375, one best sequence is selected based on the "concatenation cost"*).

8. Regarding claims 9-10, Hunt et al. further disclose the method of claim 1 further comprising forming a representation of a plurality of pronunciations of the target utterance, and wherein searching the graph includes matching any of the pronunciations of the target utterance to paths through the graph (*sections 2.1-2.2 on pages 374-375, "forced aligning"*), and forming a representation of the pronunciation of the target utterance in terms of alternating unit labels and transitions labels (*sections 2.1-2.2 on pages 374-375, concatenation of units*).

Claim Rejections - 35 USC § 103

9. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

10. Claim 4 is rejected under 35 U.S.C. 103(a) as being unpatentable over Hunt et al. (IEEE Publication).

11. Regarding claim 4, Hunt et al. disclose a method for selecting acoustic units in a concatenative speech synthesis system using Viterbi search algorithm, but fail to specifically disclose that the step of determining the best scoring path involves using a dynamic programming algorithm. However, examiner takes official notice that dynamic programming is well known in the art. The advantage using dynamic programming is to improve execution speed.

12. Claims 11-16 are rejected under 35 U.S.C. 103(a) as being unpatentable over Hunt et al. (IEEE Publication) in view of Beutnagel et al. (applicant's admitted prior art, incorporated by reference).

13. Regarding claim 11, Hunt et al. fail to specifically disclose the method of claim 1 wherein the graph includes a first part that encodes a sequence of segments and a corresponding sequence of unit labels for each of the source utterances, and a second part that encodes allowable transitions between segments of different source utterances and encodes a transition score for each of those transitions; and matching the pronunciation of the target utterance to paths through the graph includes considering paths in which each transition between segments of different source utterances

identified by that path corresponds to a different sub-path of that path that passes through the second part of the graph.

However, Beutnagel et al. teach the graph including a first part that encodes a sequence of segments and a corresponding sequence of unit labels for each of the source utterances, and a second part that encodes allowable transitions between segments of different source utterances and encodes a transition score for each of those transitions (*sections 4.1-4.3, pre-computing and caching all the possible joint costs*); and matching the pronunciation of the target utterance to paths through the graph includes considering paths in which each transition between segments of different source utterances identified by that path corresponds to a different sub-path of that path that passes through the second part of the graph (*sections 4.1-4.3, pre-computing and caching all the possible joint costs for use at runtime to reduce computing time*).

Since Hunt et al. and Beutnagel et al. are analogous art because they are from the same field of endeavors, it would have been obvious to one of ordinary skill in the art at the time of invention to modify Hunt et al. by incorporating the teaching of Beutnagel et al. in order to reduce search time at runtime to improve system's speed.

14. Regarding claims 12-13, Hunt et al. fail to specifically disclose the method of claim 11, wherein selecting the segments for synthesis includes evaluating a score for each of the considered paths that is based on the transition scores associated with the sub-paths through the second part of the graph, and wherein a size of the second part of the graph is substantially independent of a size of the source corpus, and a

complexity of matching the pronunciation through the graph grows less than linearly with the size of the corpus. However, Beutnagel et al. teach the step of selecting the segments for synthesis includes evaluating a score for each of the considered paths that is based on the transition scores associated with the sub-paths through the second part of the graph (*sections 4.1-4.3*), and wherein a size of the second part of the graph is substantially independent of a size of the source corpus, and a complexity of matching the pronunciation through the graph grows less than linearly with the size of the corpus (*sections 4.1-4.3, pre-computed and cached possible joint costs, units are available for used by the speech synthesis system*).

Since Hunt et al. and Beutnagel et al. are analogous art because they are from the same field of endeavors, it would have been obvious to one of ordinary skill in the art at the time of invention to modify Hunt et al. by incorporating the teaching of Beutnagel et al. in order to reduce search time at runtime to improve system's speed.

15. Regarding claim 14, Hunt et al. fail to specifically disclose the method of claim 1 further comprising: providing the corpus of source utterances, each source utterance being segmented into a sequence of segments, each consecutive pair of segments in a source utterance forming a segment boundary, and each speech segment being associated with a unit label and each segment boundary being associated with a transition label; and forming the graph, including forming a first part of the graph that encodes a sequence of segments and a corresponding sequence of unit labels for each of the source utterances, and forming a second part that encodes allowable transitions

between segments of different source utterances and encodes a transition score for each of those transitions.

However, Beutnagel et al. teach the steps of providing the corpus of source utterances, each source utterance being segmented into a sequence of segments, each consecutive pair of segments in a source utterance forming a segment boundary, and each speech segment being associated with a unit label and each segment boundary being associated with a transition label (*sections 4.1-4.3, pre-computed and cached possible joint costs, units are available for used by the speech synthesis system*); and forming the graph, including forming a first part of the graph that encodes a sequence of segments and a corresponding sequence of unit labels for each of the source utterances, and forming a second part that encodes allowable transitions between segments of different source utterances and encodes a transition score for each of those transitions (*sections 4.1-4.3, pre-computed and cached possible joint costs, units are available for used by the speech synthesis system*).

Since Hunt et al. and Beutnagel et al. are analogous art because they are from the same field of endeavors, it would have been obvious to one of ordinary skill in the art at the time of invention to modify Hunt et al. by incorporating the teaching of Beutnagel et al. in order to reduce search time at runtime to improve system's speed.

16. Regarding claim 15, Hunt et al. further disclose the method of claim 14 wherein forming the second part of the graph is performed independently of the utterances in the corpus of source utterances (*can be speaker independent models*).

17. Regarding claim 16, Hunt et al. fail to specifically disclose the method of claim 14 further comprising: augmenting the corpus of source utterances with additional utterances; and augmenting the graph including augmenting the first part of the graph to encode the additional utterances, and linking the augmented first part to the second part without modifying the second part based on the additional utterances. However, Beutnagel et al. teach the step of augmenting the corpus of source utterances with additional utterances (*sections 4.1-4.3, Viterbi algorithm searches the graph and picks best units and path*); and augmenting the graph including augmenting the first part of the graph to encode the additional utterances, and linking the augmented first part to the second part without modifying the second part based on the additional utterances (*sections 4.1-4.3, pre-computing and caching all possible join costs for used by the speech synthesis system*).

Since Hunt et al. and Beutnagel et al. are analogous art because they are from the same field of endeavors, it would have been obvious to one of ordinary skill in the art at the time of invention to modify Hunt et al. by incorporating the teaching of Beutnagel et al. in order to reduce search time at runtime to improve system's speed.

18. Claim 17 is rejected under 35 U.S.C. 103(a) as being unpatentable over Hunt et al. (IEEE Publication) in view of Mohri et al. (US 6243679).

19. Regarding claim 17, Hunt et al. do not disclose that the graph is associated with a finite-state transducer which accepts input symbols that include unit labels and transition labels, and that produces identifiers of segments of the source utterances, and wherein searching the graph is equivalent to composing a finite-state transducer representation of a pronunciation of the target utterance with the finite-state transducer with which the graph is associated.

However, Mohri et al. teach that the graph is associated with a finite-state transducer which accepts input symbols that include unit labels and transition labels, and that produces identifiers of segments of the source utterances (col. 10, ln. 28 to col. 11, ln. 67), and wherein searching the graph is equivalent to composing a finite-state transducer representation of a pronunciation of the target utterance with the finite-state transducer with which the graph is associated (col. 11, ln. 31-67).

Since Hunt et al. and Mohri et al. are analogous ad because they are from the same field of endeavors, it would have been obvious to one of ordinary skill in the art at the time of invention to modify Hunt et al. by incorporating the teaching of Mohri et al. in order to achieve time and space minimization efficiencies (col. 1, ln. 60 to col. 2, ln. 2).

Conclusion

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Huyen X. Vo whose telephone number is 571-272-7631. The examiner can normally be reached on M-F, 9-5:30.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Wayne Young can be reached on 571-272-7582. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

HXV 7/19/2005

***

SUSAN MCFADDEN
PRIMARY EXAMINER